

Beyond LSH: General Binary Context for Atlas Neural Networks

One way to think about Atlas neural networks is that they accept two complementary forms of input:

- an analog feature vector
- a binary context vector

In the original formulation, the binary context consists of locality-sensitive hash (LSH) bits computed from the input vector. These bits provide a geometric partitioning of the input space, allowing different linear maps to become active in different regions.

However, there is no fundamental requirement that these binary context bits originate from LSH.

Instead, they can be any collection of binary features whose statistics make them useful conditioning variables.

Examples include:

- LSH bits
- binarized text tokens
- binary linguistic features
- grammar indicators
- object-presence flags
- metadata bits
- learned binary codes
- sparse retrieval indicators
- application-specific binary predicates

The important consideration is statistical quality rather than origin. Good binary context variables should ideally have properties such as:

- reasonably balanced activation frequencies
- useful information content
- low unnecessary redundancy
- meaningful correlation with the prediction task
- enough independence to create many distinguishable contexts

For text-to-image generation, one could imagine separating the representation into two complementary streams.

The binary stream could consist of binarized token indicators or other discrete semantic features extracted from the text.

The analog stream could be obtained by projecting the text embedding into a lower-dimensional floating-point representation using a random projection or another inexpensive projection method. Random projections are attractive because they approximately preserve distances while greatly reducing dimensionality, allowing the analog pathway to retain coarse semantic geometry.

The network would then receive:

- Binary context:
 - token presence
 - token groups
 - semantic indicator bits
 - other discrete features
- Analog vector:
 - projected text embedding
 - compressed continuous semantic representation

This creates an interesting division of labor.

The binary context selects which local linear computation should be active, while the analog vector supplies the continuous information processed within that selected region.

Viewed this way, Atlas becomes less a particular LSH architecture and more a general framework for combining continuous representations with discrete conditioning variables. LSH is simply one particularly elegant method because its binary features have a clear geometric interpretation and naturally partition the input space, but other sources of binary context may be equally appropriate for different modalities and applications.